Coordinate Descent

Nipun Batra

July 26, 2025

IIT Gandhinagar

Coordinate Descent

- Another optimisation method (akin to gradient descent)

Coordinate Descent

- Another optimisation method (akin to gradient descent)
- Objective: $\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$

Coordinate Descent

- Another optimisation method (akin to gradient descent)
- Objective: $\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$
- Key idea: Sometimes difficult to find minimum for all coordinates

Coordinate Descent

- Another optimisation method (akin to gradient descent)
- Objective: $\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$
- Key idea: Sometimes difficult to find minimum for all coordinates
- ..., but, easy for each coordinate

Coordinate Descent

- Another optimisation method (akin to gradient descent)
- Objective: $\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$
- Key idea: Sometimes difficult to find minimum for all coordinates
- ..., but, easy for each coordinate
- turns into a one-dimensional optimisation problem

Coordinate Descent

- Picking next coordinate:

Coordinate Descent

- Picking next coordinate:

Coordinate Descent

- Picking next coordinate: random, round-robin
- No step-size to choose!

Coordinate Descent

- Picking next coordinate: random, round-robin
- No step-size to choose!
- Converges for Lasso objective

Coordinate Descent : Example

Learn $y = \theta_0 + \theta_1 x$ on following dataset, using coordinate descent where initially $(\theta_0, \theta_1) = (2, 3)$ for 2 iterations.

x	y
1	1
2	2
3	3

Coordinate Descent : Example

Our predictor, $\hat{y} = \theta_0 + \theta_1 x$

Error for i^{th} datapoint, $\epsilon_i = y_i - \hat{y}_i$

$\epsilon_1 = 1 - \theta_0 - \theta_1$

$\epsilon_2 = 2 - \theta_0 - 2\theta_1$

$\epsilon_3 = 3 - \theta_0 - 3\theta_1$

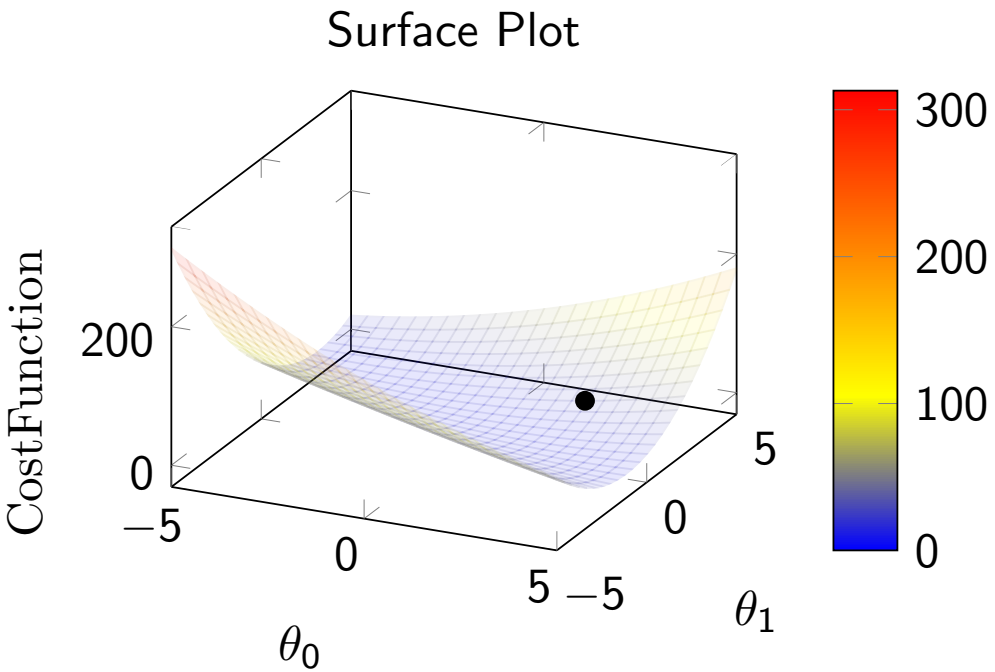
$\text{MSE} = \frac{\epsilon_1^2+\epsilon_2^2+\epsilon_3^2}{3} = \frac{14+3\theta_0^2+14\theta_1^2-12\theta_0-28\theta_1+12\theta_0\theta_1}{3}$

Iteration 0

$$\text{MSE} = \frac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$

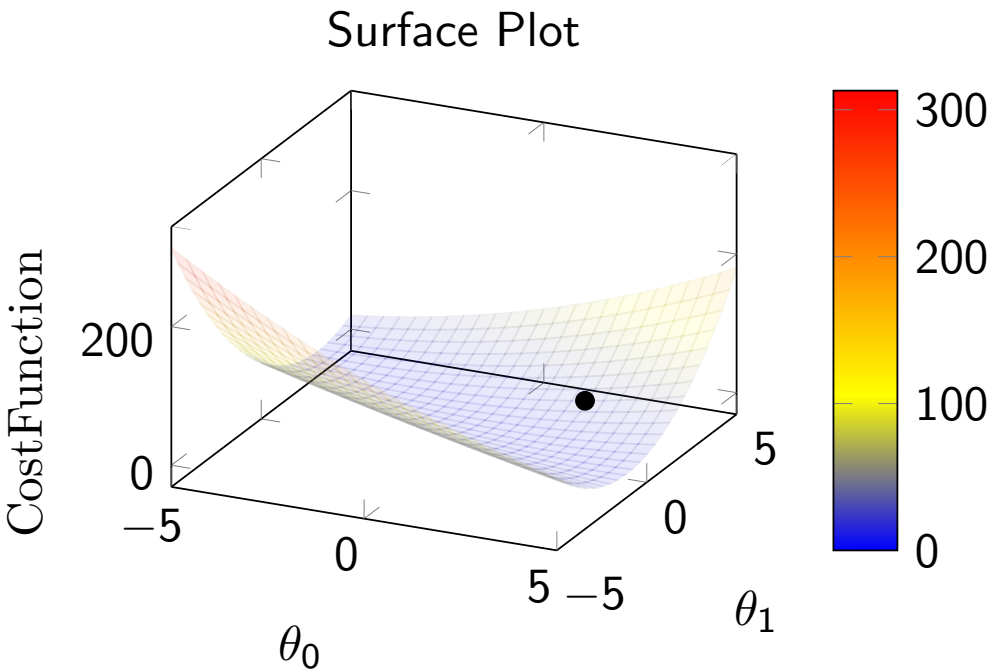
Iteration 0

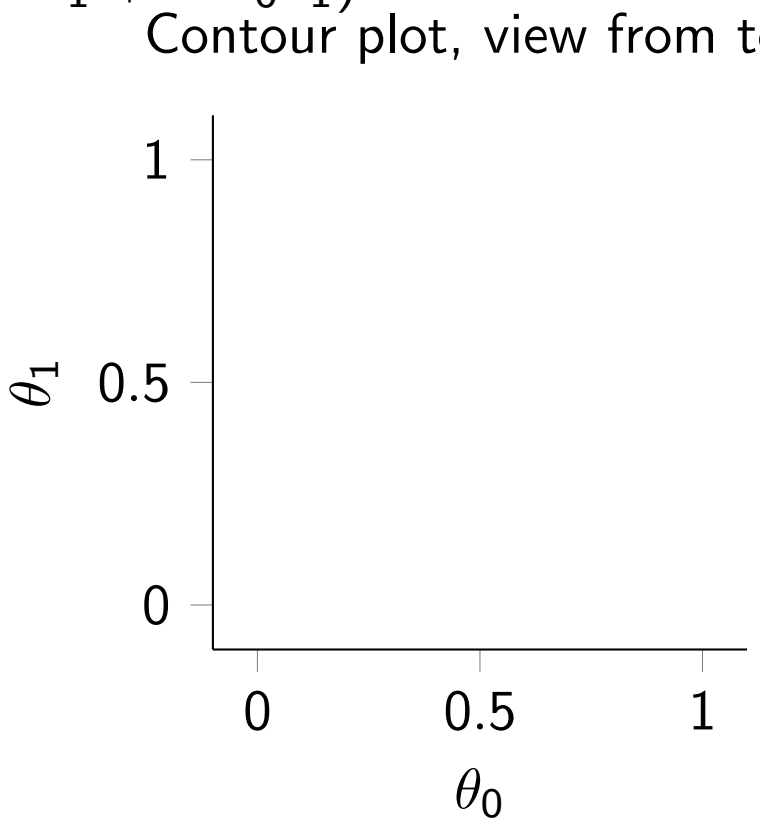
$$\text{MSE} = \tfrac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$



Iteration 0

$$\text{MSE} = \tfrac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$





Coordinate Descent : Example

Iteration 1

INIT: $\theta_0 = 2$ and $\theta_1 = 3$

$\theta_1 = 3$ optimize for θ_0

Coordinate Descent : Example

Iteration 1

INIT: $\theta_0 = 2$ and $\theta_1 = 3$

$\theta_1 = 3$ optimize for θ_0

$\frac{\partial \text{MSE}}{\partial \theta_0} = 6\theta_0 + 24 = 0$

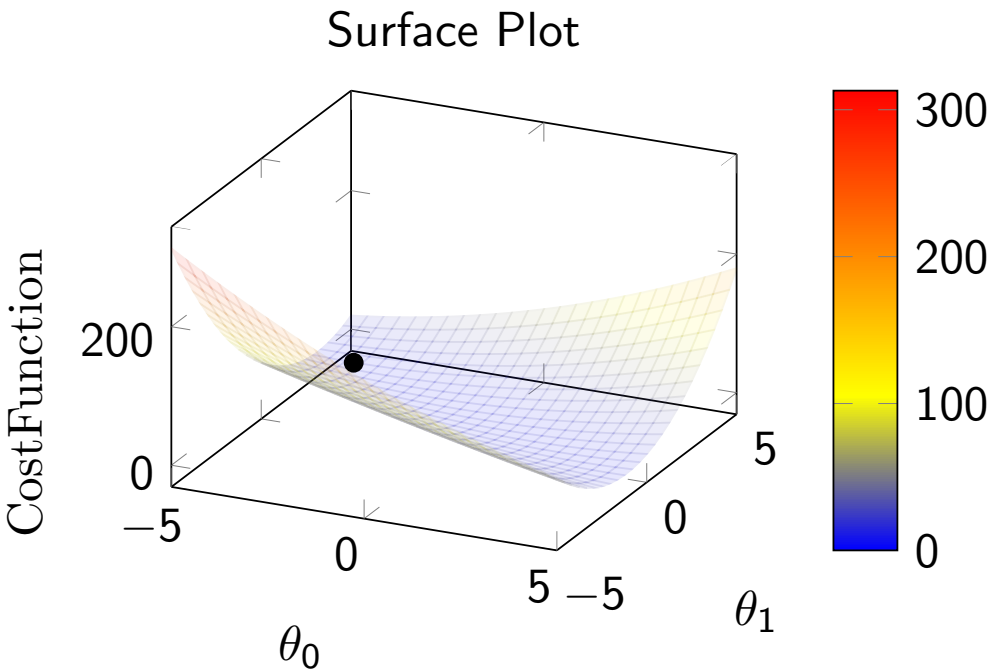
$\theta_0 = -4$

Iteration 1

$$\text{MSE} = \frac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$

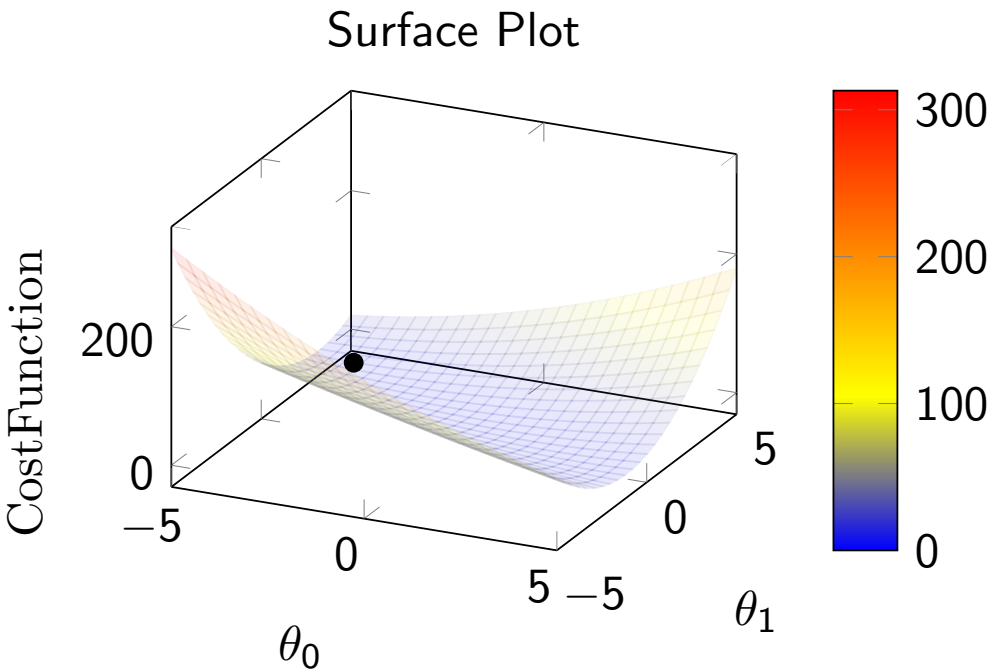
Iteration 1

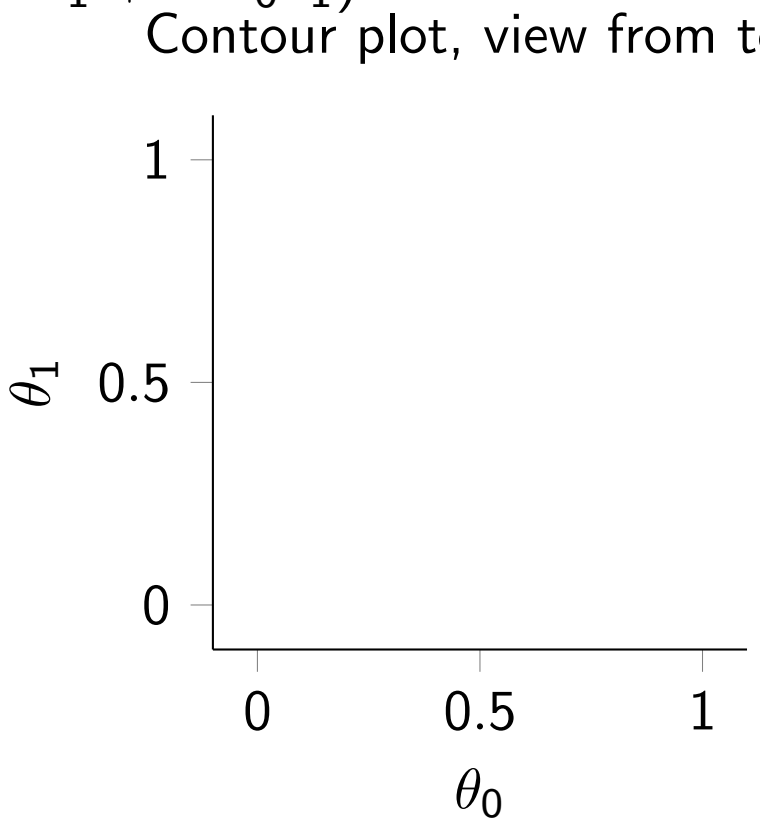
$$\text{MSE} = \frac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$



Iteration 1

$$\text{MSE} = \tfrac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$





Coordinate Descent : Example

Iteration 2

INIT: $\theta_0 = -4$ and $\theta_1 = 3$

$\theta_0 = -4$ optimize for θ_1

Coordinate Descent : Example

Iteration 2

INIT: $\theta_0 = -4$ and $\theta_1 = 3$

$\theta_0 = -4$ optimize for θ_1

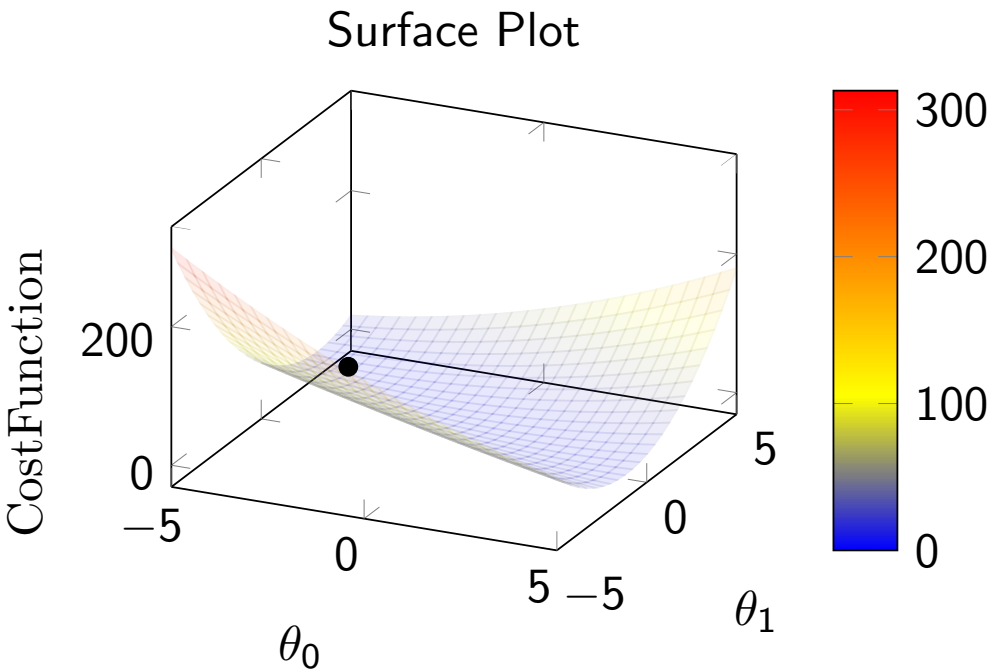
$\theta_1 = 2.7$

Iteration 2

$$\text{MSE} = \frac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$

Iteration 2

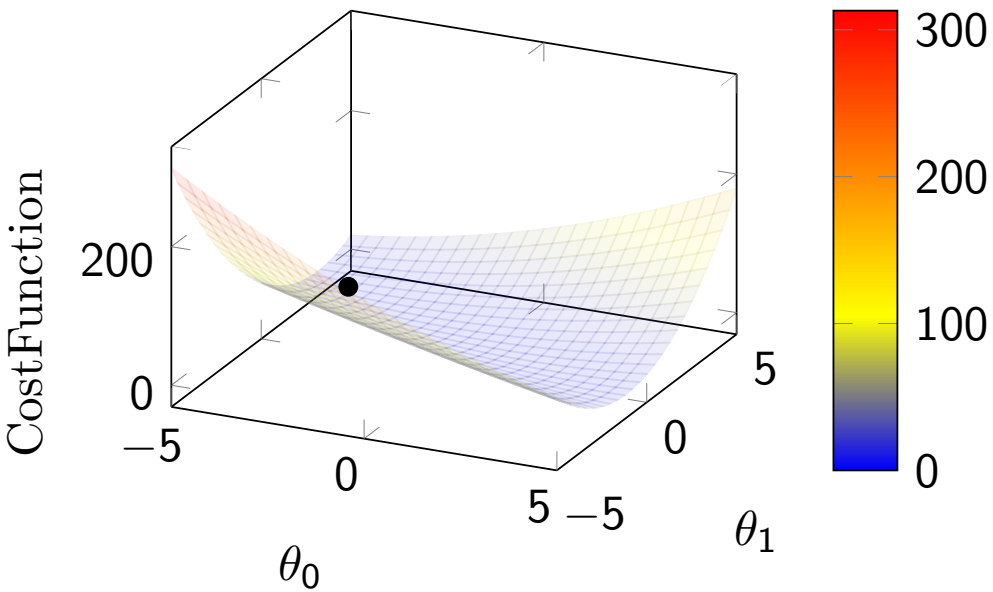
$$\text{MSE} = \frac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$



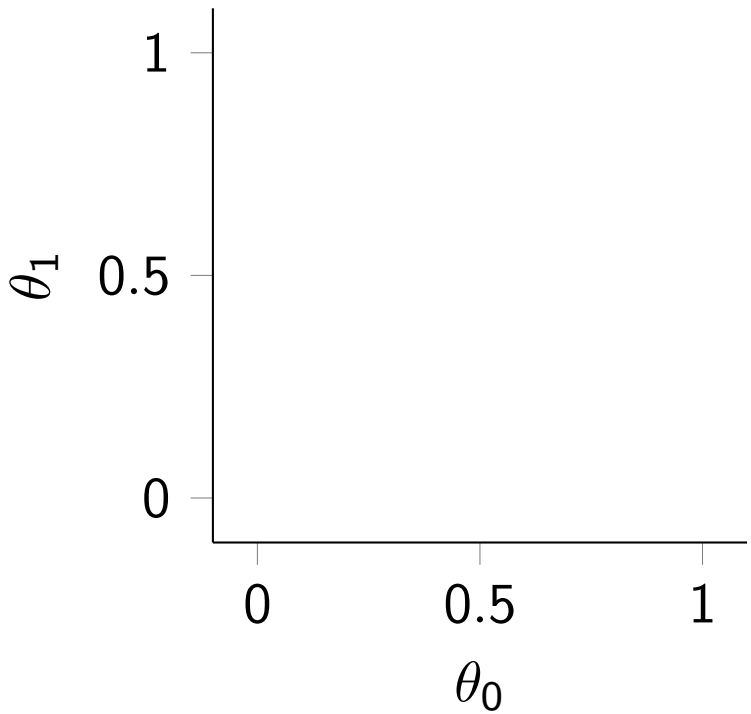
Iteration 2

$$\text{MSE} = \tfrac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$

Surface Plot



Contour plot, view from top



Coordinate Descent : Example

Iteration 3

INIT: $\theta_0 = -4$ and $\theta_1 = 2.7$

$\theta_1 = 2.7$ optimize for θ_0

Coordinate Descent : Example

Iteration 3

INIT: $\theta_0 = -4$ and $\theta_1 = 2.7$

$\theta_1 = 2.7$ optimize for θ_0

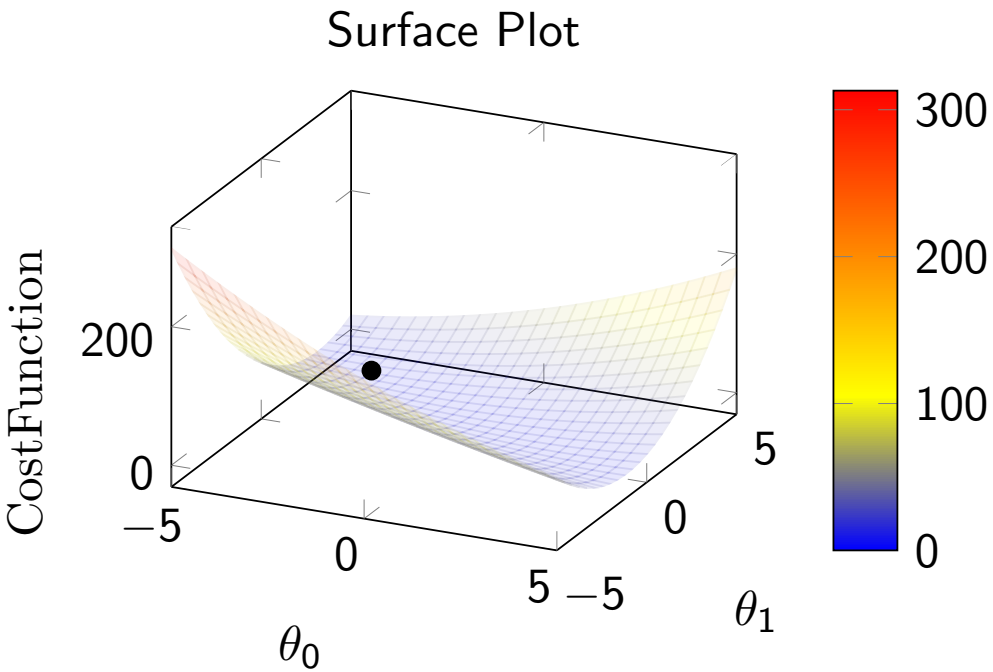
$\theta_0 = -3.4$

Iteration 3

$$\text{MSE} = \frac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$

Iteration 3

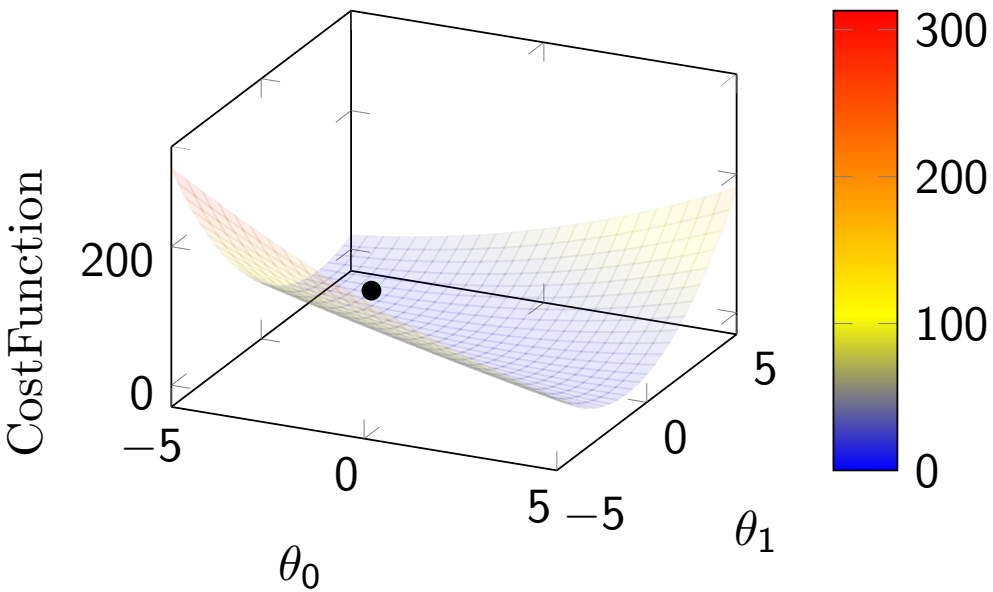
$$\text{MSE} = \frac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$



Iteration 3

$$\text{MSE} = \tfrac{1}{3}(14 + 3\theta_0^2 + 14\theta_1^2 - 12\theta_0 - 28\theta_1 + 12\theta_0\theta_1)$$

Surface Plot



Contour plot, view from top